2010-01-04

# Many Sequence-Specific Chromatin Modifying Protein-Binding Motifs Show Strong Positional Preferences for Potential Regulatory Regions in the Saccharomyces Cerevisiae Genome

*1772–1779* *Nucleic Acids Research, 2010, Vol. 38, No. 6* *Published online 4 January 2010*
*doi:10.1093/nar/gkp1195*

# Many sequence-specific chromatin modifying protein-binding motifs show strong positional preferences for potential regulatory regions in the *Saccharomyces cerevisiae* genome

**Loren Hansen[1,2], Leonardo Mariño-Ramírez[3] and David Landsman[1,*]**

[1]Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8900 Rockville Pike, Bethesda, MD 20894, [2]Bioinformatics Program, Boston University, Boston, MA 02215, USA and [3]Computational Biology and Bioinformatics Unit, Plant Molecular Genetics Laboratory, Biotechnology and Bioindustry Center, Corporacion Colombiana de Investigacion Agropecuaria – CORPOICA Bogota, Colombia

Received October 9, 2009; Revised December 7, 2009; Accepted December 8, 2009

## ABSTRACT

**Initiation and regulation of gene expression is critically dependent on the binding of transcriptional regulators, which is often temporal and position specific. Many transcriptional regulators recognize and bind specific DNA motifs. The length and degeneracy of these motifs results in their frequent occurrence within the genome, with only a small subset serving as actual binding sites. By occupying potential binding sites, nucleosome placement can specify which sequence motif is available for DNA-binding regulatory factors. Therefore, the specification of nucleosome placement to allow access to transcriptional regulators whenever and wherever required is critical. We show that many DNA-binding motifs in *Saccharomyces cerevisiae* show a strong positional preference to occur only in potential regulatory regions. Furthermore, using gene ontology enrichment tools, we demonstrate that proteins with binding motifs that show the strongest positional preference also have a tendency to have chromatin-modifying properties and functions. This suggests that some DNA-binding proteins may depend on the distribution of their binding motifs across the genome to assist in the determination of specificity. Since many of these DNA-binding proteins have chromatin remodeling properties, they can alter the local nucleosome structure to a more permissive and/or restrictive state, thereby assisting in determining DNA-binding protein specificity.**

## INTRODUCTION

At any given point in time, cells are performing complex programs of gene expression. The binding of transcriptional regulators to target genes determines their expression or repression. Many DNA-binding proteins (DBPs) recognize and bind specific DNA sequence motifs located within specific regulatory regions of the gene. However, the length and nucleic acid composition of these binding motifs frequently enables their random occurrence within the genome, sometimes up to thousands of repetitions. Therefore, sequence information alone is insufficient to completely determine specificity (1,2).

Within the nucleus, DNA exists in complexes with RNA and proteins called chromatin. Commonly composed of an octamer of histone proteins consisting of two copies each of histones H2A, H2B, H3 and H4, nucleosomes are the basic repeating units of chromatin [for review see ref. (3)]. DNA wraps around the histone octamer core in approximately two superhelical turns. These cores are spaced ~10–80 bp apart; this internucleosomal DNA is referred to as linker DNA. This DNA can vary in length significantly, even between neighboring nucleosomes. DNA within nucleosomes is less accessible to DBPs, including transcriptional regulators (4). It has long been thought that by occupying potential binding sites, nucleosomes play an indirect role in regulating gene expression (4–7). However, this raises the question of how the structure of chromatin is constructed initially to ensure the availability of sites for transcriptional regulator binding. It is likely that inherent signals within the DNA sequence play an important role in positioning nucleosomes (8,9). Also critical are chromatin remodeling factors (CRFs) that reposition or modify nucleosomes (8,10–13), thereby repressing or

*To whom correspondence should be addressed. Tel: +1 301 435 5981; Fax: +1 301 480 2288; Email: landsman@ncbi.nlm.nih.gov

enhancing transcription. Whether and how CRFs act to modify chromatin structure to a more permissive/restrictive state remains unknown. One possibility is that CRFs rely on the quality and genomic position of their DNA sequence motifs to help establish specificity. In this study, we investigated this hypothesis by examining the positional distribution of predicted binding sites for 184 DBPs in the *Saccharomyces cerevisiae* genome.

## MATERIALS AND METHODS

### Calculating promoter enrichment scores

Transcription start sites (TSSs), as well as promoter and coding sequences, were obtained from the UCSC genome browser (14). The Mining Yeast Binding Sites (MYBS) database was used to obtain 666 position weight matrices (PWMs) (15). The Spt10 PWM was obtained from ref. (16) for a total of 667 PWMs. Promoters were defined as regions extending 1000 bp upstream of TSSs, excluding any coding sequence. Each PWM was used to score both promoter and coding sequences while looking for subsequences that closely match the binding motif represented by the PWM. The score of each subsequence was derived from the sum of the position-specific score of each nucleotide composing the subsequence. For a subsequence of length $l(s_1 \ldots s_l)$ with length $l$ equal to the number of columns in the PWM, the score was calculated as

$$\text{Score} = \sum_{j=1}^{l} m_{s_j,j} \quad \mathbf{1}$$

where $S_j$ represents the nucleotide at position $j$ of subsequence $s$ and $m_{i,j}$ represents the score in the PWM for row $i$ and column $j$.

We randomized the sequence of interest by shuffling the nucleotides while retaining the overall nucleotide composition. Then each set of randomized sequence was scanned against the set of PWMs and the number of high-scoring matches was counted. The randomization was performed 800 times, and the mean and standard deviation for the number of matches expected in the randomized sequence for a given PWM was calculated. A $z$-score representing the degree of sequence motif enrichment was calculated using

$$Z = \frac{x - u_r}{\sigma_r}, \quad \mathbf{2}$$

where $x$ is the number of high-scoring matches for the unshuffled sequence, $u_r$ is the mean number of high-scoring matches for 800 sets of shuffled sequences, and $\sigma_r$ is the standard deviation for the group of 800 sets of shuffled sequences. We then used the calculated $z$-scores from the promoter and coding sequence to calculate a promoter enrichment score (i.e. promoter $z$-score − ORF $z$-score) for each PWM.

To perform this analysis, it was necessary to select a cutoff score. Therefore, similar to other comparable studies (17), a cutoff score representing 70% of the maximum possible score for a given PWM was chosen. Results from analyses using cutoff scores representing 80 and 90% of the maximum possible score showed little differences.

### Gene ontology (GO) analysis

The set of PWMs was filtered using the methods outlined below and then ranked according to the promoter preference score. Finally, using the online David GO tool, we searched for enriched GO terms (18) in the top 20% of PWMs ($N = 37$). As a control, we assessed the set of all proteins (184) represented by the collection of 667 PWMs used in this study. To avoid the use of an arbitrary percentile cutoff, we also applied the online Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) tool (19) to our set of ranked proteins. GOrilla uses a flexible threshold technique to search for GO terms enriched in a ranked list.

The set of PWMs used contained considerable redundancy (i.e. many DBPs are associated with multiple PWMs). To perform the GO analysis, it was necessary to filter the set of 667 PWMs to obtain a unique set of 184 PWMs to pair with the 184 unique proteins. Two different filtering methods were used to determine which PWM out of the set of PWMs associated with a given DBP would be used when ranking the protein. With the first method, we filtered PWMs based on the promoter enrichment score. The PWM with the highest promoter enrichment score from the set of PWMs was selected to pair with that protein. Each protein was then ranked according to the promoter enrichment score of its paired PWM and GO analysis performed as outlined above. Using this method, both analysis tools identified GO terms related to chromatin modification for the highly ranked proteins. With the second method, we filtered the PWMs according to information content. The PWM with the highest information content was selected to pair with its associated protein. We repeated the above analysis using both GO tools. Using the David tool, we again identified an enrichment of chromatin modifying GO terms for highly ranked proteins ($P<0.05$). However, GOrilla did not reveal any GO terms possibly due to the stringent cutoff ($P<0.001$) of this tool.

### Nucleosome overlap score

With the set of high-scoring matches in promoter regions and a map of nucleosome positions produced in a recent study (20), we calculated the fraction of predicted binding sites that overlapped with a well-positioned nucleosome for each PWM. Nucleosomes, unlike many DBPs, do not necessarily have a well-defined binding site. Instead, they may have multiple binding locations in different cells for the same nucleosome. For each nucleosome, Mavrich *et al.* (20) calculated a 'fuzziness score' that represented the extent a nucleosome varies its binding location. To obtain a list of well-positioned nucleosomes we ranked all nucleosomes by their fuzziness score and took the top 15%.

To calculate the significance of the observed overlap of predicted binding sites with well-positioned nucleosomes, we randomly changed the positions of the predicted binding sites within a 1000-bp window and calculated

the fraction of randomized sites that overlapped with a well-positioned nucleosome. After 1000 iterations, the mean and standard deviation of nucleosome overlap were estimated. In addition, a concurrent $z$-score representing the degree of nucleosome overlap above or below random chance was calculated according to Equation (2), where $x$ was the fraction of high-scoring matches that overlapped a nucleosome, $u_r$ was the mean fraction of high-scoring matches that overlap a nucleosome calculated based on 1000 random permutations, and $\sigma_r$ represented the standard deviation of the fractional overlap of the randomly moved high-scoring matches.

Promoter regions have a tendency to contain nucleosome-depleted regions (21). To control for potential bias, we randomly changed the predicted binding site location within a 1000-bp window that was centered on the binding site. In doing so, the randomly permuted binding sites were still mostly positioned within the same local chromatin structure. A 1000-bp window will almost always include some of the neighboring ORF sequences. Thus, while restricting the randomization to a defined window reduces the effect of simply being within a promoter region, it does not eliminate it entirely. One could argue that our results indicating a strong bias toward promoter regions for some motifs exacerbate this issue. However, in our calculation of nucleosome occupancy, we only used those sites found in promoter regions. Hence, promoter bias should not play a significant role in these analyses. For each PWM we paired its promoter enrichment score with its nucleosome overlap score and calculated the correlation using Spearman rank correlation. Correlation coefficients were calculated using those PWMs with at least 50 predicted binding sites.

### Within promoter positional analysis

For each high-scoring promoter region match, we calculated the distance to the closest TSS. Predicted binding sites that could not be mapped to a TSS were discarded. Sequence motifs that were highly 'location constrained' within promoter regions clustered together. For every PWM that had at least 50 predicted binding sites within promoter regions, we obtained the distance from the TSS for every high-scoring match (i.e. predicted binding site) and then calculated the mean, median and semi-interquartile range for the distance distribution. The smaller the semi-interquartile range, the more clustered the predicted binding sites were and the stronger the location constraint within promoter regions.

## RESULTS

### Many DBP sequence motifs displayed strong preferences for promoter regions as opposed to coding regions

Sequence motifs for DBPs are commonly represented by a position weight matrix (PWM) (1,22). We obtained a set of 667 PWMs representing binding motifs for 184 DBPs from the MYBS database (15). For each PWM we calculated a promoter enrichment score. The larger the score, the more enriched the sequence motif was in promoter regions relative to coding regions.

Not surprisingly, most sequence motifs showed dramatically greater enrichment in promoter regions than in coding regions (Figure 1). For example, Orc1p, which has been demonstrated to function in chromatin modification (23), displayed the greatest difference in enrichment between promoter and coding sequence. For this sequence motif, the number of high-scoring matches within the promoter region was 1240, corresponding to a $z$-score of 261. Meanwhile, the number of high-scoring sequence motif matches within coding sequence was 38, corresponding to a $z$-score of $-0.88$. Yeast contains $\sim$8.4 Mb of coding sequence compared to $\sim$2.5 Mb of promoter sequence. Despite this, the Orc1p motif occurred far more often in potential regulatory, but not coding, sequence in the yeast genome.

### Sequence motifs showing a strong positional preference were also enriched for CRFs

We then investigated whether proteins whose sequence motifs showed a high positional preference for promoter regions also shared common biological functions. To explore this question, the set of 184 proteins was sorted according to the promoter enrichment score from largest to smallest ('Materials and methods' section). Then the online David bioinformatics resource tool (http://david.abcc.ncifcrf.gov/home.jsp) (18) was used to assess GO terms associated with the top 20% of ranked proteins. Chromatin remodeling-related terms were highly represented among these highly ranked proteins ($P < 0.05$), including chromatin modification, establishment and/or maintenance of chromatin architecture, DNA packaging, gene silencing, negative regulation of gene expression epigenetic, chromatin silencing and heterochromatin formation.

To verify these results, we performed a similar analysis using the GOrilla tool (http://cbl-gorilla.cs.technion.ac.il/) (19). When given a ranked list of genes, GOrilla searches for GO terms that show greater enrichment for items near the top of the list relative to the rest of the list. Therefore, it was unnecessary to limit this analysis to the top 20% of ranked proteins. We submitted to GOrilla a set of proteins ranked according to their promoter enrichment score and examined GO term enrichment. Similar to the analysis using David, many chromatin-associated GO terms were identified for high-ranking proteins, including histone modification, covalent chromatin modification, and chromatin modification. This analysis indicates that DBPs whose sequence motifs showed the strongest positional constraint for promoters were also associated with CRFs.

### A negative correlation exists between high positional preference and nucleosome occupancy

The relationship revealed above between the positional preference of sequence motifs and CRFs led us to postulate that a correlation may also exist between the binding of proteins exhibiting a high positional preference and nucleosome occupancy. Based on nucleosome positions

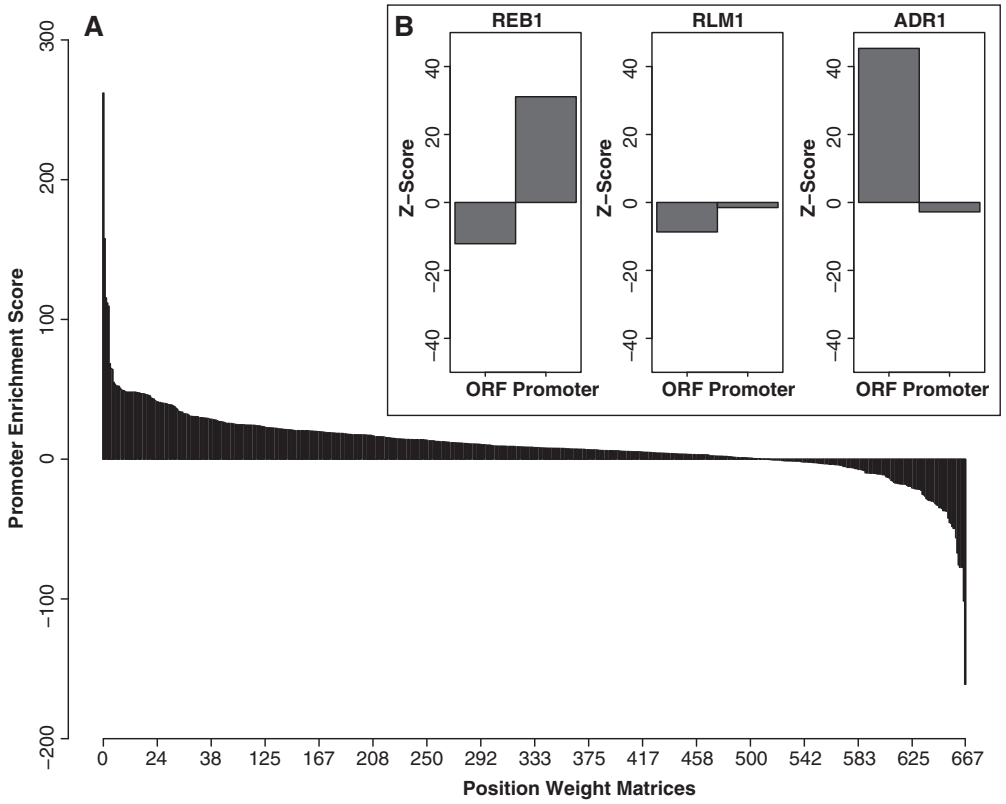


**Figure 1.** Promoter enrichment scores for all 667 PWMs. (**A**) The promoter enrichment score for all 667 PWMs sorted by decreasing promoter enrichment. (**B**) Examples of specific PWMs taken from the top, middle and end of the set of PWMs sorted by promoter enrichment. From left to right, the first graph is a random example from the top 50 PWMs (REB1). The second graph (RLM1) is a random example selected from PWMs ranked 300–400. The third graph (ADR1) is a random example taken from PWMs ranked 600–667. Plotted in (B) is the *z*-score for each PWM indicating over- or under-representation in the given class of sequence elements. A positive *z*-score denotes over-representation while a negative *z*-score signifies under-representation.

obtained in a recent Chip-Seq study (20), we calculated a score to represent nucleosome occupancy (see 'Materials and methods' section) for each PWM.

A large negative score indicated that the overlap between predicted binding sites and nucleosomes was much less than would be expected by random chance. Conversely, a large positive score suggested that the likelihood of an overlap was greater than random chance. The Spearman rank correlation coefficient between the promoter enrichment score and the score representing nucleosome occupancy of predicted binding sites was then calculated. Indeed, there was a negative correlation between positional preference and nucleosome occupancy ($r_s = -0.39$, $P < 1e{-}16$) (Figure 2A). The *P*-values for correlation coefficients were calculated according to Best and Roberts (24). This result, combined with those from the GO analysis, suggests that DBPs whose binding sites show strong positional preference may act in part to remove or shift nucleosomes upon binding to allow entry by other transcriptional regulators (10), thereby playing a role in determining specificity.

To further confirm these results, we repeated the correlation analysis using a different measure of nucleosome occupancy. Kaplan *et al.* (8) produced a high-resolution map of nucleosome occupancy across the yeast genome. For each position in the genome, a nucleosome occupancy score was calculated. A negative number indicated that nucleosome occupancy was below the genome average, while a positive number represented an above average likelihood for occupancy. We obtained the data set from Kaplan *et al.* (8) and averaged the nucleosome occupancy score for the set of predicted binding sites in promoter regions for a given PWM. Then, the Spearman rank correlation between the promoter enrichment score and the average nucleosome occupancy was calculated. With this method, we again observed a correlation between nucleosome occupancy and promoter preference ($r_s = -0.44$, $P < 1e{-}16$) (Figure 2B).

Kaplan *et al.* also produced a map of nucleosome occupancy for chromatin that was reconstituted *in vitro*. Our results suggest that the trend toward lower nucleosome occupancy for motifs with a high positional preference may be due to active chromatin remodeling by the transcription factors that bind those motifs. As such, we would expect to observe a positive correlation between positional preference and those motifs that showed the largest difference between *in vitro* and *in vivo* nucleosome occupancy. To test this hypothesis, we calculated the correlation between the promoter enrichment score and the difference in nucleosome occupancy *in vitro* and *in vivo* for the set of predicted binding sites in promoter regions for each PWM. As anticipated, promoter enrichment and the

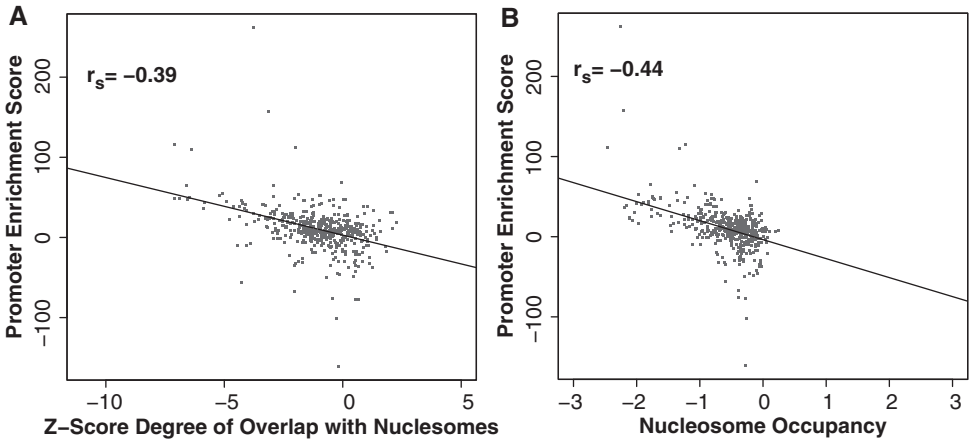


**Figure 2.** Scatter plots showing the correlation between promoter enrichment and nucleosome occupancy for predicted sites in promoter regions. The *y*-axis represents promoter enrichment (the larger the value, the stronger the positional preference for promoter regions). The *x*-axis is a score representing the degree of nucleosome occupancy. For each PWM with at least 50 predicted binding sties in promoter regions, the promoter enrichment score was plotted against the degree of nucleosome occupancy for predicted binding sites using that PWM. (**A**) A scatter plot using the overlap of predicted transcription factor binding sites with well-positioned nucleosomes as the measure of nucleosome occupancy. (**B**) The method for measuring nucleosome occupancy as described by Kaplan *et al.* (8).

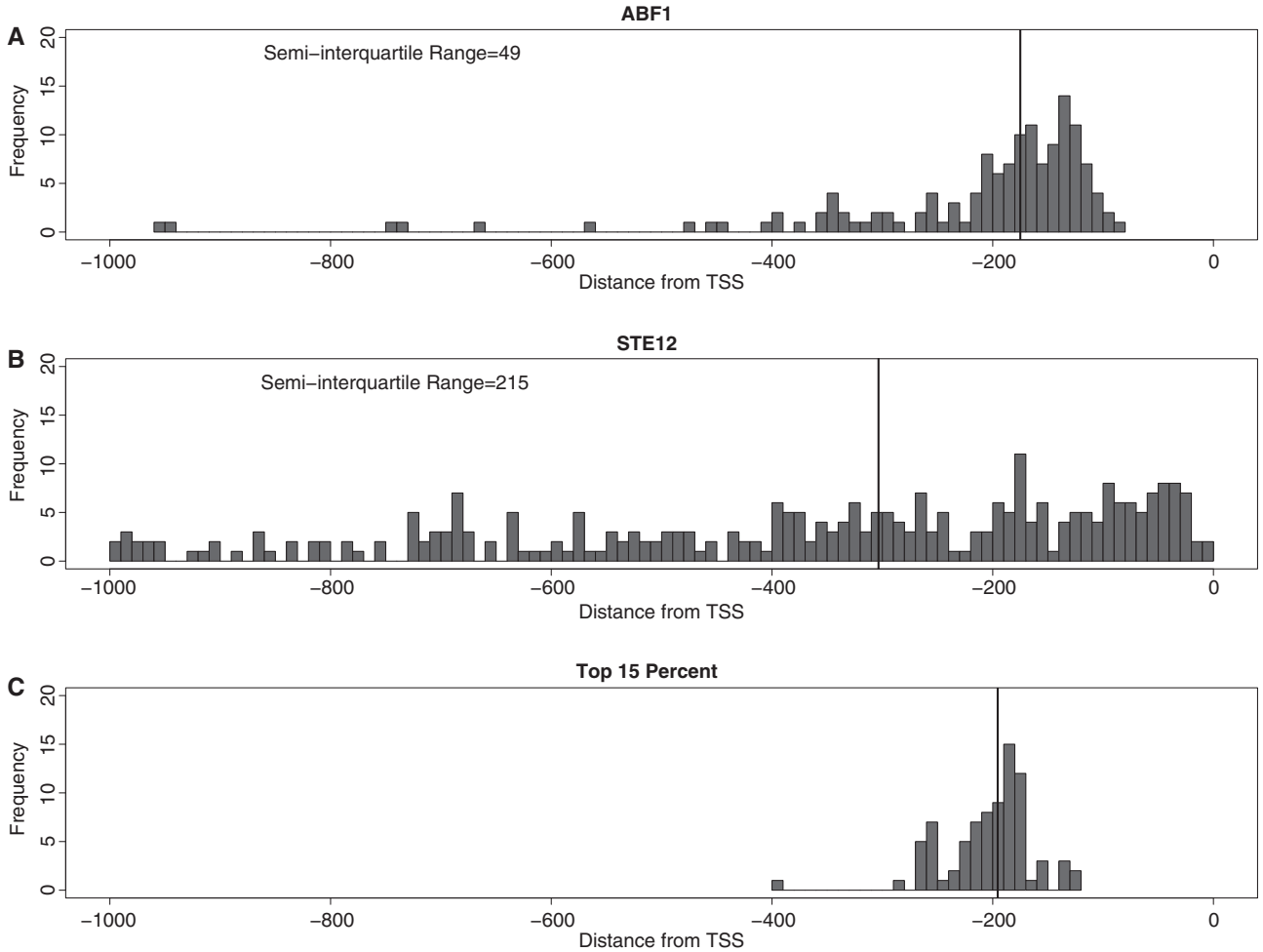


**Figure 3.** Distribution of high-scoring matches in promoter regions. (**A**) Example of a sequence motif (ABF1) that is location constrained within promoter regions. Plotted are histograms of distances from the TSS for all predicted binding sites for the indicated DNA-binding protein. The TSS is marked by position zero and the black line represents the median (*x*-axis units are in bps). (**B**) Example of a sequence motif that is not location constrained within promoter regions (STE12). (**C**) All sequence motifs that had at least 50 predicted binding sites were ranked on the basis of semi-interquartile range. Plotted are the medians of the distance distributions for the top 15% ($N = 83$), representing the sequence motifs with a strong promoter positional bias.

difference between *in vitro* and *in vivo* nucleosome occupancy was positively correlated ($r_s = 0.46$, $P < 1e{-}16$, see Supplementary Figure 1).

### A correlation exists between high promoter enrichment and strong location constraint within promoters

Previous studies have shown that motif context, including distance from the TSS, likely plays a role in gene regulation in yeast and humans (25,26). This prompted us to investigate whether sequence motifs showing strong promoter enrichment also display a strong positional constraint within promoter regions. To answer this question, we calculated the distance to the TSS for predicted binding sites in yeast promoters. Sequence motifs that demonstrated significant location constraint within promoter regions clustered together at similar distances from the TSS corresponding to a narrow distribution of distances (Figure 3A). Sequence motifs that were not constrained within the promoter exhibited distance distributions with

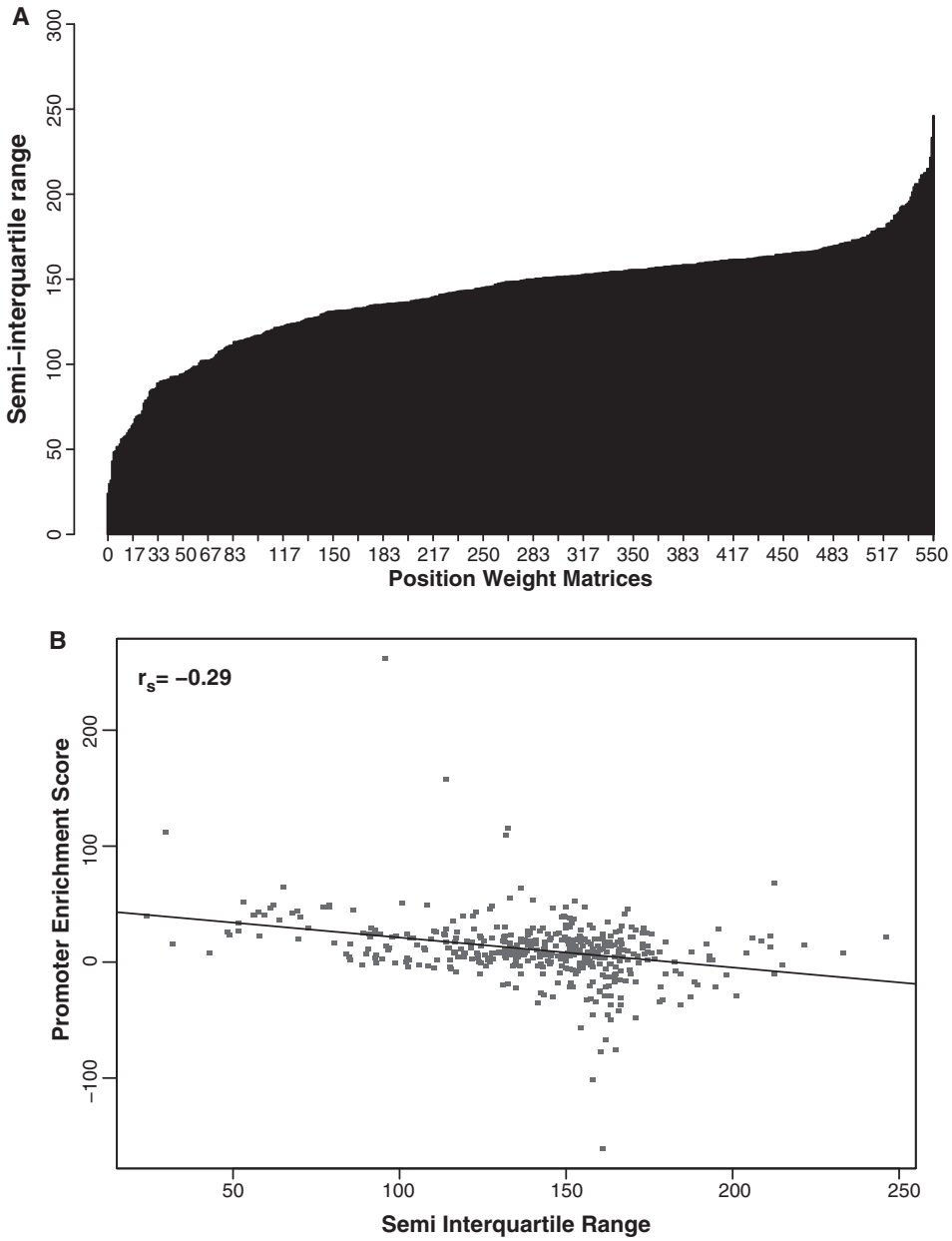


**Figure 4.** Plot of positional preference within promoter regions. (**A**) The semi-interquartile range for the set of PWMs with at least 50 predicted binding sites within promoter regions ($N = 551$) sorted by increasing semi-interquartile range. The semi-interquartile range measured the distribution dispersion. The larger the value, the greater the distribution spread. A smaller semi-interquartile range indicates more location constraint within promoter regions for the predicted binding sites. (**B**) A scatter plot depicting the correlation between promoter enrichment and positional preference within promoter regions. The *y*-axis represents promoter enrichment in which larger values signify greater enrichment in promoter regions relative to coding regions. The *x*-axis represents the degree of positional preference within promoter regions. The smaller the value the more clustered the predicted binding sites are within promoter regions, and the higher the degree of positional preference within promoter regions.

a larger spread (Figure 3B). We noticed with interest that sequence motifs with a strong positional bias within promoter regions seem to cluster ~100–300 bp upstream of the TSS (Figure 3C).

The semi-interquartile range was calculated to measure the distribution spread statistically. Because many of the distance distributions were skewed (see Figure 3a), the semi-interquartile range was a better measure of spread than standard deviation. The Spearman rank correlation coefficient between the positional preference score and the semi-interquartile range was calculated. Indeed, a correlation between positional preference for promoter regions (high promoter enrichment) and positional preference within promoter regions ($r_s = -0.29$, $P = 2.4e{-}12$) (Figure 4) was revealed.

## DISCUSSION

Recent work elucidating nucleosome positioning in yeast has revealed a common chromatin architecture around TSS's consisting of a nucleosome covering the TSS, an immediate upstream nucleosome-free region (NFR) of ~140 bp, and a well-positioned nucleosome ('−1' nucleosome) on the upstream border of the NFR (7,27). Veners *et al*. (28) demonstrated that the −1 nucleosome is evicted upon recruitment of RNA polymerase II. Additionally they showed that a number of chromatin remodeling complexes were selectively associated with the −1 nucleosome. Furthermore, a number of sequence-specific experimentally determined binding sites overlapped the −1 nucleosome. These results support the idea that the positioning of the −1 nucleosome may be strongly regulated.

Here we show that sequence motifs with a strong positional bias within promoter regions cluster almost exclusively ~100–300 bp upstream of the TSS (Figure 3C). This localization places them in a prime location to regulate or be regulated by the −1 nucleosome, further supporting the idea that positioning of the −1 nucleosome is important in transcriptional regulation.

If CRFs with sequence motifs that exhibit strong positional preferences are modifying the chromatin structure in part to provide specificity to other DBPs, what is the mechanism of action? One possibility is that CRFs remove and/or shift nucleosomes to open up binding sites for other transcriptional regulators. For example, Rap1p, Abf1p and Reb1p are all highly abundant sequence-specific general regulatory factors that bind motifs with a strong preference for promoter regions. There is good evidence that all three play a role in influencing chromatin structure (10,29,30). Additionally, these proteins appear to act in part by creating bubbles of open chromatin (8,31–33). In the case of Rap1p and Abf1p, creating a region of open chromatin appears to facilitate the binding of additional regulatory factors, leading to transcription enhancement (31). In many cases, Rap1p and Abf1p are unable to activate robust transcription alone (34,35) and require additional regulatory factors. Further support is provided by the observation that Rap1p- and Abf1p-binding sites can be substituted for one another without a loss in function (31,35).

However, both Rap1p and Abf1p are involved in many functions, including repression (36–38). Rap1p initiates a repressive chromatin structure by interacting directly with the chromatin modifying factors Sir3p and Sir4p (37). Therefore, in addition to making binding sites accessible, it is likely that DBPs whose sequence motifs show a strong positional preference can increase specificity by directly interacting with chromatin modifiers or transcriptional regulators.

A question that immediately presents itself is whether or not the pronounced preference for promoter regions is sufficient to determine specificity. Is the positional distribution sufficient to fully explain binding *in vivo*? In a genome-wide location analysis, Lieb *et al*. (39) noted the strongly skewed positional preference of Rap1p-binding motifs and concluded that the positional distribution of potential Rap1p-binding sites may account for much of the specificity in Rap1p binding. However, the skewed positional distribution of these potential binding sites was insufficient in fully explaining the pattern of Rap1p binding. For the case of Rap1p, additional genome-wide mechanisms also appear to be at work.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank John Spouge for reviewing the manuscript and for his suggestions.

## FUNDING

L.M.R. was supported by CORPOICA. Funding for open access charge: Intramural Program of the National Library of Medicine, National Institutes of Health.

*Conflict of interest statement*. None declared.

## REFERENCES

1. D'Haeseleer,P. (2006) What are DNA sequence motifs? *Nat. Biotechnol.*, **24**, 423–425.
2. Ren,B., Robert,F., Wyrick,J.J., Aparicio,O., Jennings,E.G., Simon,I., Zeitlinger,J., Schreiber,J., Hannett,N., Kanin,E. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
3. Felsenfeld,G. and Groudine,M. (2003) Controlling the double helix. *Nature*, **421**, 448–45.
4. Liu,X., Lee,C.K., Granek,J.A., Clarke,N.D. and Lieb,J.D. (2006) Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. *Genome Res.*, **16**, 1517–1528.
5. Almer,A., Rudolph,H., Hinnen,A. and Horz,W. (1986) Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *Embo J.*, **5**, 2689–2696.
6. Sekinger,E.A., Moqtaderi,Z. and Struhl,K. (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell*, **18**, 735–748.

7. Shivaswamy,S., Bhinge,A., Zhao,Y., Jones,S., Hirst,M. and Iyer,V.R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.*, **6**, e65.
8. Kaplan,N., Moore,I.K., Fondufe-Mittendorf,Y., Gossett,A.J., Tillo,D., Field,Y., Leproust,E.M., Hughes,T.R., Lieb,J.D., Widom,J. *et al.* (2008) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, **458**, 362–366.
9. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thastrom,A., Field,Y., Moore,I.K., Wang,J.P. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
10. Yu,L. and Morse,R.H. (1999) Chromatin opening and transactivator potentiation by RAP1 in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **19**, 5279–5288.
11. Whitehouse,I., Rando,O.J., Delrow,J. and Tsukiyama,T. (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, **450**, 1031–1035.
12. Morse,R.H. (2003) Getting into chromatin: how do transcription factors get past the histones? *Biochem. Cell Biol.*, **81**, 101–112.
13. Hartley,P.D. and Madhani,H.D. (2009) Mechanisms that specify promoter nucleosome location and identity. *Cell*, **137**, 445–458.
14. Karolchik,D., Kuhn,R.M., Baertsch,R., Barber,G.P., Clawson,H., Diekhans,M., Giardine,B., Harte,R.A., Hinrichs,A.S., Hsu,F. *et al.* (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.*, **36**, D773–D779.
15. Tsai,H.K., Chou,M.Y., Shih,C.H., Huang,G.T., Chang,T.H. and Li,W.H. (2007) MYBS: a comprehensive web server for mining transcription factor binding sites in yeast. *Nucleic Acids Res.*, **35**, W221–W226.
16. Eriksson,P.R., Mendiratta,G., McLaughlin,N.B., Wolfsberg,T.G., Marino-Ramirez,L., Pompa,T.A., Jainerin,M., Landsman,D., Shen,C.H. and Clark,D.J. (2005) Global regulation by the yeast Spt10 protein is mediated through chromatin structure and the histone upstream activating sequence elements. *Mol. Cell Biol.*, **25**, 9127–9137.
17. Harbison,C.T., Gordon,D.B., Lee,T.I., Rinaldi,N.J., Macisaac,K.D., Danford,T.W., Hannett,N.M., Tagne,J.B., Reynolds,D.B., Yoo,J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.
18. Dennis,G. Jr, Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, P3.
19. Eden,E., Lipson,D., Yogev,S. and Yakhini,Z. (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.*, **3**, e39.
20. Mavrich,T.N., Ioshikhes,I.P., Venters,B.J., Jiang,C., Tomsho,L.P., Qi,J., Schuster,S.C., Albert,I. and Pugh,B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, **18**, 1073–1083.
21. Yuan,G.C., Liu,Y.J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**, 626–630.
22. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
23. Triolo,T. and Sternglanz,R. (1996) Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. *Nature*, **381**, 251–253.
24. Best,D.J. and Roberts,D.E. (1975) Algorithm AS 89: the upper tail probabilities of Spearman's Rho. *J. Royal Stat. Soc. C*, **24**, 377–379.
25. Westholm,J.O., Xu,F., Ronne,H. and Komorowski,J. (2008) Genome-scale study of the importance of binding site context for transcription factor binding and gene regulation. *BMC Bioinformatics*, **9**, 484.
26. Tharakaraman,K., Bodenreider,O., Landsman,D., Spouge,J.L. and Marino-Ramirez,L. (2008) The biological function of some human transcription factor binding motifs varies with position relative to the transcription start site. *Nucleic Acids Res.*, **36**, 2777–2786.
27. Lee,W., Tillo,D., Bray,N., Morse,R.H., Davis,R.W., Hughes,T.R. and Nislow,C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.*, **39**, 1235–1244.
28. Venters,B.J. and Pugh,B.F. (2009) A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res.*, **19**, 360–371.
29. Lascaris,R.F., Groot,E., Hoen,P.B., Mager,W.H. and Planta,R.J. (2000) Different roles for abf1p and a T-rich promoter element in nucleosome organization of the yeast RPS28A gene. *Nucleic Acids Res.*, **28**, 1390–1396.
30. Chasman,D.I., Lue,N.F., Buchman,A.R., LaPointe,J.W., Lorch,Y. and Kornberg,R.D. (1990) A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. *Genes Dev.*, **4**, 503–514.
31. Yarragudi,A., Miyake,T., Li,R. and Morse,R.H. (2004) Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **24**, 9152–9164.
32. Badis,G., Chan,E.T., van Bakel,H., Pena-Castillo,L., Tillo,D., Tsui,K., Carlson,C.D., Gossett,A.J., Hasinoff,M.J., Warren,C.L. *et al.* (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell*, **32**, 878–887.
33. Angermayr,M., Oechsner,U. and Bandlow,W. (2003) Reb1p-dependent DNA bending effects nucleosome positioning and constitutive transcription at the yeast profilin promoter. *J. Biol. Chem.*, **278**, 17918–17926.
34. Devlin,C., Tice-Baldwin,K., Shore,D. and Arndt,K.T. (1991) RAP1 is required for BAS1/BAS2- and GCN4-dependent transcription of the yeast HIS4 gene. *Mol. Cell Biol.*, **11**, 3642–3651.
35. Goncalves,P.M., Griffioen,G., Minnee,R., Bosma,M., Kraakman,L.S., Mager,W.H. and Planta,R.J. (1995) Transcription activation of yeast ribosomal protein genes requires additional elements apart from binding sites for Abf1p or Rap1p. *Nucleic Acids Res.*, **23**, 1475–1480.
36. Shore,D. (1994) RAP1: a protean regulator in yeast. *Trends Genet.*, **10**, 408–412.
37. Moretti,P. and Shore,D. (2001) Multiple interactions in Sir protein recruitment by Rap1p at silencers and telomeres in yeast. *Mol. Cell Biol.*, **21**, 8082–8094.
38. Loo,S., Laurenson,P., Foss,M., Dillin,A. and Rine,J. (1995) Roles of ABF1, NPL3, and YCL54 in silencing in *Saccharomyces cerevisiae*. *Genetics*, **141**, 889–902.
39. Lieb,J.D., Liu,X., Botstein,D. and Brown,P.O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.*, **28**, 327–334.